N88-23378

PUPIL MEASURES OF ALERTNESS AND MENTAL LOAD

Richard W. Backs and Larry C. Walrath
Human Performance Laboratory
McDonnell Douglas Astronautics Company
St. Louis, Missouri

1.0 OVERVIEW

As part of our internal research and development program at McDonnell Douglas we are examining human factors engineering issues associated with how operators extract information from visual displays. Recently, we have been using psychophysiological measures of operator performance, in addition to behavioral performance measures, in order to better assess operator mental workload (MWL) associated with using a particular display configuration during performance of a visual search task.

In our work, we take a rather broad view of the concept of MWL. That is, we consider MWL to be the cognitive effort associated with performing an information processing task analogous to the physical effort required to perform a manual task. The problem with such a definition, of course, is specifying precisely what is meant by "cognitive effort." We assume that cognitive effort is determined by the <u>extent</u> to which the information processing resources required to perform the criterion task are <u>actively</u> engaged in task performance. This definition presupposes that the task can be performed within the limitations of the available resources. Unfortunately, in practice MWL very often becomes synonymous with the paradigms with which it is manipulated (such as primary and secondary tasks) or the dependent variables with which it is measured (such as behavioral performance decrements in reaction time and error rate).

Clearly, there are many determinants of MWL. Two of these are the nature of the task and the required behavioral performance. Another determinant is the capability of individual operators to allocate their processing resources in ways to efficiently and effectively perform the task. This ability to optimally allocate resources requires a combination of the operator's natural abilities, training, and motivation. Any time there is a mismatch between the optimum level of resource allocation required by the task and the optimum level at which the operator is able to engage the necessary resources, an unacceptable amount of MWL will result. This mismatch may occur either because the task requires too much or too little cognitive effort.

Further, MWL is a closed-loop process, and as such is also determined by the costs to the operator (in physiological terms) of maintaining performance. These costs are increased in tasks that require either more or less than the operator's optimal level of cognitive effort. The physiological costs become part of a feedback loop, along with the knowledge of results of the behavioral performance, and they serve as additional inputs upon the operator.

The MWL itself is, just as clearly, not a unitary phenomenon. Inappropriate load on the operator may occur at any of a number of points in the information processing flow. Although we are not testing psychological theory, we make heuristic use of several theoretical models. The first is that total information processing capacity is divided into multiple resource pools according to sensory input channels (e.g., ref. 1). The second is that information processing occurs serially, progressing through well-defined stages that can be manipulated independently (ref. 2). We further assume that, with the exception of those stages requiring access to common resources that must be shared, information processing can progress independently within each resource pool and in parallel with similar ongoing stages in other resource pools (ref. 3). We recognize that overall task performance is determined by the number and priority of sensory input channels required by a task and the amount common resource time-sharing required for task completion. However, up to this point in our research, we have not been concerned with concurrently manipulating multiple sensory channel resource pools or with the competition between pools for common resources.

We believe that in order to accurately assess an operator's MWL it is necessary to measure as many of its facets as possible. Monitoring behavioral performance is absolutely necessary since this measure is the end product to be maximized. Subjective reports of MWL can be helpful to define which elements of a task operators have trouble with. Subjective reports may also indicate circumstances in which objective measures fail to reflect deficiencies in workload and thus more sensitive objective measures are required. In our research, we use psychophysiological measures to provide such a sensitive measure. An added benefit is that the psychophysiological measures serve as a window into how the operator is allocating resources. Our goal is to discover which external (task) determinants contribute to MWL and which internal (cognitive) processes are inappropriately loaded.

As an example of our progress toward assessing operator MWL during visual search, we will present data from a recent study measuring evoked pupillary responses and response time to search displays that varied with regard to their density, use of color coding, and type of information abstraction required to complete the search. This study consisted of a single task, and was one of a series of studies originally designed to evaluate the effects of different display parameters on search time. It is meant to serve as an illustration of how adding psychophysiological response measures can help localize points of mental overload.

In a previous study (ref. 4), we described how eye-movement analysis was used to determine the effects of information density, use of color coding, and type of information abstraction on visual display search time. In that study, we found that search time and the number of fixations required to search a display increased with the density of the display. Longer search times and more fixations were also required to count the number of target items in a display than to locate a single target. However, even though search time was longer for monochrome than for color-coded displays, the number of fixations required to search these displays did not differ. Instead, the duration of each fixation was shorter for color-coded than for monochrome displays indicating that subjects processed symbolic information more efficiently using a color code than using a shape code.

We also obtained evoked pupillary responses in reference 4 in order to evaluate this measure as an indicator of information processing load (e.g., refs. 5 and 6). Single-trial pupillary responses observed in reference 4 had a distinctive tri-phasic shape (dilation-constriction-dilation) similar to the average pupillary response data reported in reference 7. Significant effects of color coding and color coding by type of information abstraction were obtained for the initial dilation-constriction phase following display onset. However, an uncontrolled change in luminance preceding the search display was subsequently discovered. That change could possibly have accounted for the unexpectedly large constriction. In the present study, the luminance problem was corrected and the basic search task was repeated on another sample of subjects. In addition, these subjects participated in a psuedo-search condition which was included as a control for nontask-related luminance and color effects of the displays.

2.0 METHOD

2.1 Subjects

Eight McDonnell Douglas Corp. employees participated as subjects. Two of the subjects were female, and the age of all subjects ranged from 19 to 42 years. One subject had participated in reference 4, and another subject had previously completed the search task; both of these subjects were placed in the group that received the active search condition first. All other subjects were naive to the experimental procedure.

2.2 Apparatus

A Data General Eclipse S-140 minicomputer was used to generate the stimulus displays, control and time the experimental events, and collect and reduce for analysis the pupil diameter and response time data. Displays were presented on an AED 512 high-resolution color graphics terminal. Pupil diameter data were collected at 60 Hz using an Applied Science Eye View Monitor and TV Pupillometer System model 1994-S. The experimental set-up is shown in Figure 1. All photometry to calibrate luminance of the stimulus displays was performed with a Photo Research Co. Spectra-Pritchard Model 1980-A photometer using a photopic filter.

2.3 Procedure

Subjects participated in two experimental sessions: an active search task (SEARCH) where they were required to abstract information from a display, and a passive psuedo-search task (CONTROL) where they received the same task as in the SEARCH condition but were not required to abstract information from a display. SEARCH and CONTROL conditions were administered on successive days. Half of the subjects (one female) received the SEARCH condition first, while the other half received the CONTROL condition first.

Subjects viewed four different displays for each combination of the Information Density (10 vs 20 symbols), Color Coding (redundant with symbol shape vs monochromatic symbology), and Search Type (COUNT vs LOCATE a specific symbol: requiring exhaustive or self-termininating search strategies, respectively) independent variables for a total of 32 trials in both the SEARCH and CONTROL conditions. The order of presentation for the 32 displays was determined randomly for each subject in both experimental sessions.

Trials consisted of a series of four screens. The first was a calibration screen with a central fixation point and four calibration points that defined the 8.8° square area of the display containing the symbology. The second was a question screen, presented for 6 sec, identifying the search type and, in the SEARCH condition, the target symbol. The target symbol was always presented in the color in which it would appear in the display (i.e., yellow rectangles, red triangles, and green semicircles for the color-coded condition or all green symbology for the noncoded condition). The third screen was the calibration screen. The display screen was presented only if subjects fixated within 1° of the central fixation point for 0.5 sec during the calibration screen. If no such fixation occurred within 2 sec, the question screen was presented again and the trial was repeated until the subject did fixate on the central point. The fourth screen was the display, which was presented only after central fixation had been verified. Figure 2 contains examples of question, calibration, and high and low density display screens.

The procedure in SEARCH and CONTROL conditions was identical except for the search and response instructions given to the subject. In the SEARCH condition, subjects actively searched the display for the target and made a button press, which terminated the display, to indicate that they had completed their search. This response time to search the display was measured in msec from display onset. Subjects then verbally reported the number of targets (for the COUNT trials) or the the quadrant of the display in which the target was located (for the LOCATE trials). Whenever subjects failed to complete a search within 6 sec, the display screen was replaced by the calibration screen and they were required to guess at the correct answer. In the CONTROL condition, subjects were not given a target to search for on the question screen; instead, they were told to merely scan each display until it terminated. Also, subjects had no responses to perform. The experimenter controlled the length of the display screen, varying it from 2-6 sec, and no verbal response was necessary.

The 32 different display screens were approximately balanced with respect to the distribution of symbols, the location of targets within the four quadrants, and the frequency of the correct answer (1, 2, 3, or 4 targets in the COUNT condition and quadrants 1-4 in the LOCATE condition). Luminance of all text and symbology on the displays was equated at 0.51 fL. Overall screen luminance within the 8.8° search area was equated for all screens (at 0.52 fL) by varying background luminance. Ambient illumination was 8.49 x 10^{-2} ft-c.

2.4 Data Quantification

Single-trial pupillary responses exhibited the characteristic tri-phasic shape previously reported (refs. 4 and 7). Figure 3 shows representative single-trial responses from a low density, color-coded trial and a high density, noncoded trial. Several measurements were made for each trial, baseline (pupil diameter at display onset) and three "components" (points of inflection for dilation or constriction). The first component (D1) was a small initial dilation that peaked about 266 msec after display onset. The second component (C) was a large constriction that peaked about 941 msec after display onset. These components were followed by a gradual dilation (D2), the resolution of which depended upon display duration. The differences between the D1 and C components and the D2 and C components were also computed for analysis. The D1-C difference was computed to determine the relative size of the constriction from the point of onset. The D2-C difference was computed to determine the amount of pupillary dilation that occurred from the point of maximum constriction. If the point of maximum dilation did not occur prior to the motor response, then the last data point in the trial was used as D2. Each of these measures and the search time were averaged over the four trials of each combination of Information Density, Color Coding, and Search Type.

All analyses were performed with the SAS General Linear Models procedure (ref. 8). A Latin square (ref. 9) was used to balance the effects of Group (SEARCH or CONTROL condition first), Condition (SEARCH or CONTROL), and Day (first or second test day), while the effects of Density, Color Coding, and Search Type were totally within-subjects. The degrees of freedom for all F ratios were (1,6) with the comparison-wise error rate set at $p < 0.05$. Duncan's Multiple Range tests were performed for all significant main effects and two-way interactions using the SAS Duncan procedure.

3.0 RESULTS

The main effect of Condition ($F = 11.52$) was significant for the baseline measure, reflecting the overall larger pupil diameter in the SEARCH than in the CONTROL condition. This effect was probably due to a generalized arousal difference between the two conditions as it was significant for all component measures. In order to correct for this initial difference, the baseline was subtracted from each component prior to analysis. Where results for component and peak-to-peak difference scores overlap, we will report only the peak-to-peak data.

The peak-to-peak difference scores, D1-C and D2-C, were both affected by the Condition and Color Coding manipulations, but in distinctly different ways. As shown in Figure 4 (left panel), the main effects of Condition ($F = 13.28$) and Color Coding ($F = 88.83$) were significant for the D1-C component, and these effects did not interact. Pupil diameter was larger overall (i.e., the size of the constriction was smaller) in the SEARCH than in the CONTROL condition, and pupil diameter was also larger for noncoded as opposed to color-coded displays. However, for D2-C (Figure 4, right panel), only the Condition by Color Coding interaction was significant ($F = 11.30$). Although none of the pair-wise comparisons differed significantly, pupil diameter for the D2-C component was larger for noncoded than for color-coded displays in the SEARCH condition, consistent with the D1-C data. However, in the CONTROL condition, pupil diameter was larger for the color-coded than for the noncoded displays.

The Condition by Search Type interaction was significant for both D1-C and D2-C ($F = 9.14$ and 18.37, respectively). The form of the interaction, however, was quite different for the two components. For the D1-C component (Figure 5, left panel), pupil diameter was larger (i.e., less constriction) in the SEARCH than in the CONTROL condition, and the difference between SEARCH and CONTROL conditions was greater in the LOCATE (self-terminating search) than in the COUNT (exhaustive search) trials. For the D2-C component (Figure 5, right panel), there was a crossover interaction in which no comparisons between means differed significantly. However, pupil diameter in the SEARCH condition was larger (i.e., greater dilation) in the COUNT than in the LOCATE trials.

The interaction between Density by Color Coding was significant for the D2 component ($F = 11.09$). As can be seen in Figure 6, pupil diameter for color-coded displays was larger for high-density than low-density displays. The opposite was found for noncoded displays, with larger pupil diameters found for the low-density displays. The difference between high- and low-density displays was not significant in either color-coding condition, however.

Search times (from the SEARCH condition) were significantly shorter for low vs high density displays ($F = 42.52$), for color-coded vs noncoded displays ($F = 34.08$), and for LOCATE vs COUNT trials ($F = 16.18$). However, the Density by Search Type ($F = 10.52$) and Color Coding by Search Type ($F = 16.54$) interactions were also significant. Search times were faster for low than for high density displays for both COUNT and LOCATE trials, but this difference was much greater for COUNT trials. Similarly, color coding decreased search time for both COUNT and LOCATE trials, but had a much greater effect for COUNT trials. The search time data for these two interactions can be seen in Figure 7.

4.0 DISCUSSION

The evoked pupillary response was sensitive to information processing demands in a visual search task. In particular, larger pupillary diameter was observed in the SEARCH condition where subjects were actively processing information relevant to task performance, as opposed to the CONTROL condition where subjects passively viewed the displays. However, the large baseline difference between the SEARCH and CONTROL conditions may only have indicated that subjects were more aroused in the active search task than in the psuedo-search task. In fact, many subjects complained of boredom and fatigue in the psuedo-search task.

Of greater import was that larger pupillary diameter, corresponding to longer search time, was observed for noncoded than for color-coded displays in the SEARCH condition. The Condition by Color Coding interaction for the D2-C difference component indicated that this effect was not an artifact of intensity differences between the color and monochrome displays or a result of the color displays having greater stimulatory value than the monochrome displays simply because they activated more photoreceptors. If pupil diameter was determined solely by some physical dimension of the displays, the same type of response would have been elicited in both the SEARCH and CONTROL conditions. Instead, pupil diameter was larger to the color displays in the CONTROL condition, presumably because they were intrinsically more interesting than the monochrome displays.

The only effect of the display density manipulation was the Density by Color Coding interaction for the D2 component. This interaction was probably due to our procedure of terminating data collection at display offset along with the motor response. This procedure could have resulted in truncating the D2 component in the low-density color-coded condition when the trial was very easy and, consequently, response time was very short. Alternatively, D2 resolution may not have been completed in some high-density noncoded trials, particularly when the trial was very difficult and subjects did not complete their search within the 6-sec limit. Because of our procedure, it was unclear precisely how display density affects the pupillary response. It is clear, however, that task difficulty (at least as manipulated by color coding) interacts with display density to determine maximal pupil dilation.

In summary, these data indicate the potential usefulness of pupillary responses in evaluating the information processing requirements of visual displays. However, because our task was originally designed to evaluate visual search behavior, and not pupillary responses, several methodological deficiencies limited the conclusions that can be drawn from the data. We are currently in the process of adapting the visual search paradigm to the examination of pupillary responses in order to conduct further research in this area. The promise of the approach lies in the separation of the impact of some of the multiple determinants of mental workload.

5.0 REFERENCES

1. Wickens, C.D., & Derrick, W.: Workload measurement and multiple resources. Proceedings, 1981 IEEE conference on cybernetics and society. New York: Institute of Electrical and Electronics Engineers, 1981.

2. Carroll, J.B.: Individual difference relations in psychometric and experimental cognitive tasks. Report No. 163. Chapel Hill, NC: Thurstone Psychometric Laboratory, 1980.

3. Wickens, C.D., Sandry, D., & Vidulich, M.: Compatibility and resource competition between modalities of input, central processing, and output: Testing a model of complex task performance. Human Factors, 25, 1983, 227-248.

4. Walrath, L.C., Backs, R.W., & Hancock, G.A.: Eye movement during visual search: Effects of information density, color coding and search type. Psychophysiology, 23, 1986, 469. (Abstract).

5. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91, 1982, 276-292.

6. Goldwater, B.C.: Psychological significance of pupillary movements. Psychological Bulletin, 77, 1972, 340-355.

7. Libby, W.L., Lacey, B.C., & Lacey, J.I.: Pupillary and cardiac activity during visual attention. Psychophysiology, 10, 1973, 270-294.

8. SAS Institute Inc.: SAS Users Guide: Statistics, 1982 Edition. Cary, NC: SAS Institute Inc, 1982.

9. Winer, B.J.: Statistical Principals in Experimental Design. New York: McGraw Hill, 1962.

Figure 1. Pupil diameter data collection.

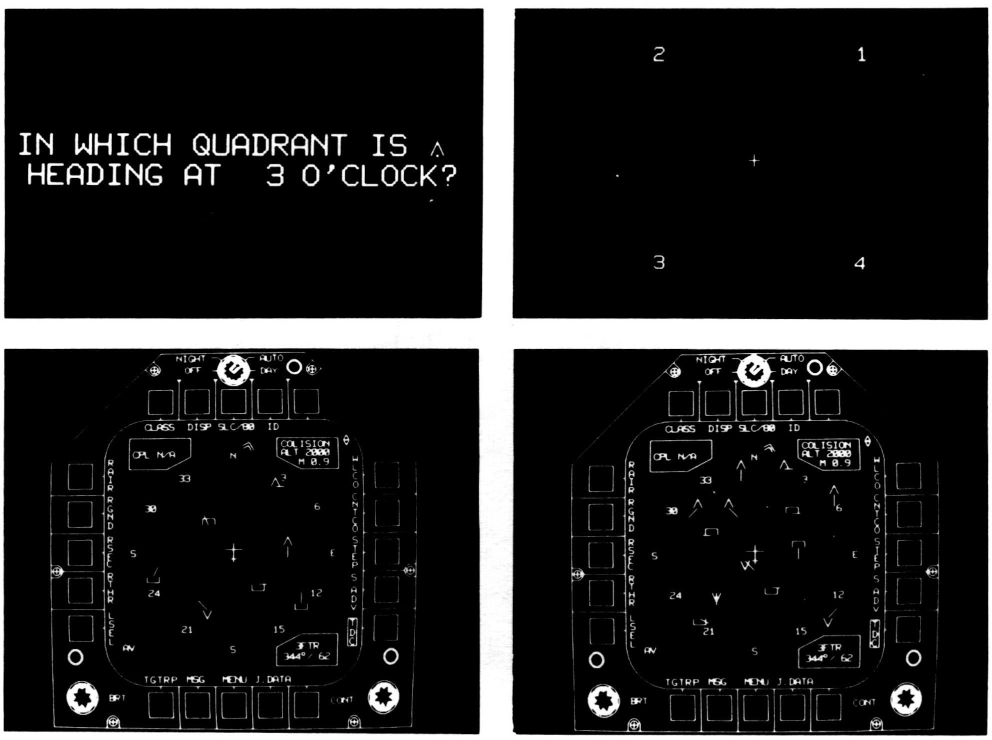

Figure 2. Examples of a question screen from the count condition (upper left), the calibration screen (upper right), a high-density display (lower right), and a low-density display (lower left).

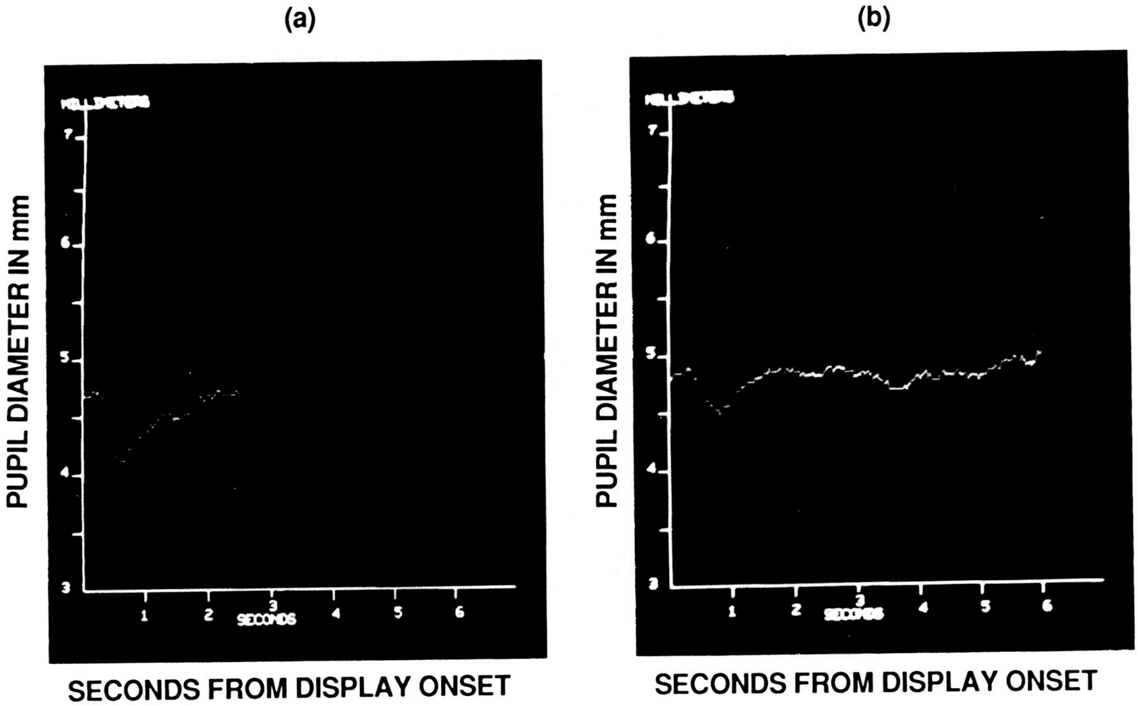

Figure 3. Illustrative single-trial pupillary responses from (a) color-coded, low-density, LOCATE and (b) noncoded, high-density, COUNT trials.

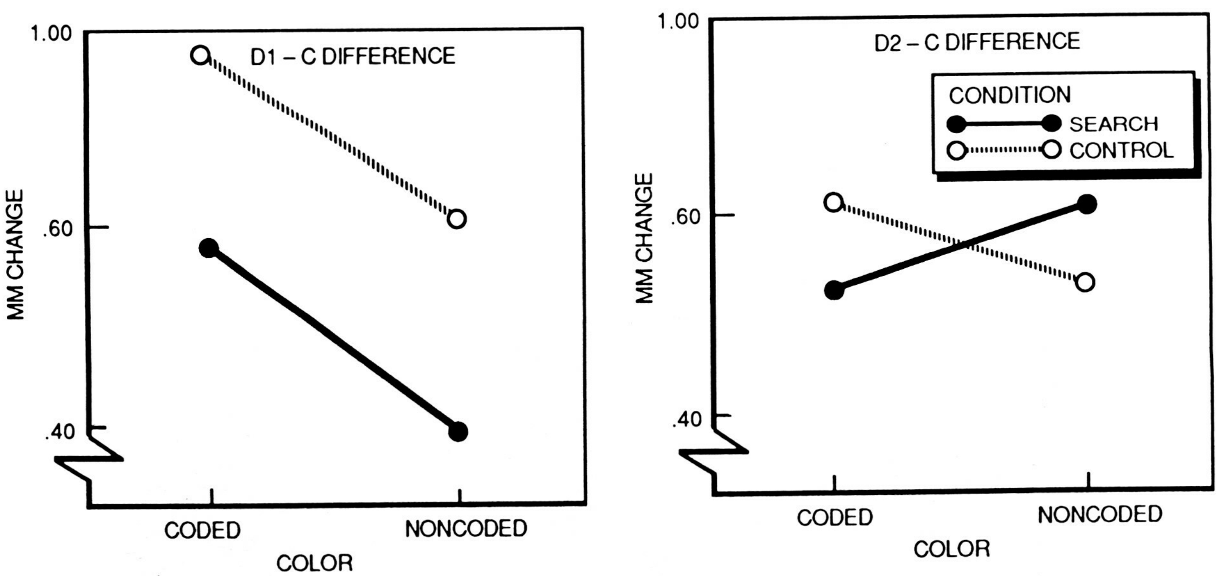

Figure 4. Color Coding and Condition effects for pupillary responses (n=8).

ORIGINAL PAGE IS
OF POOR QUALITY

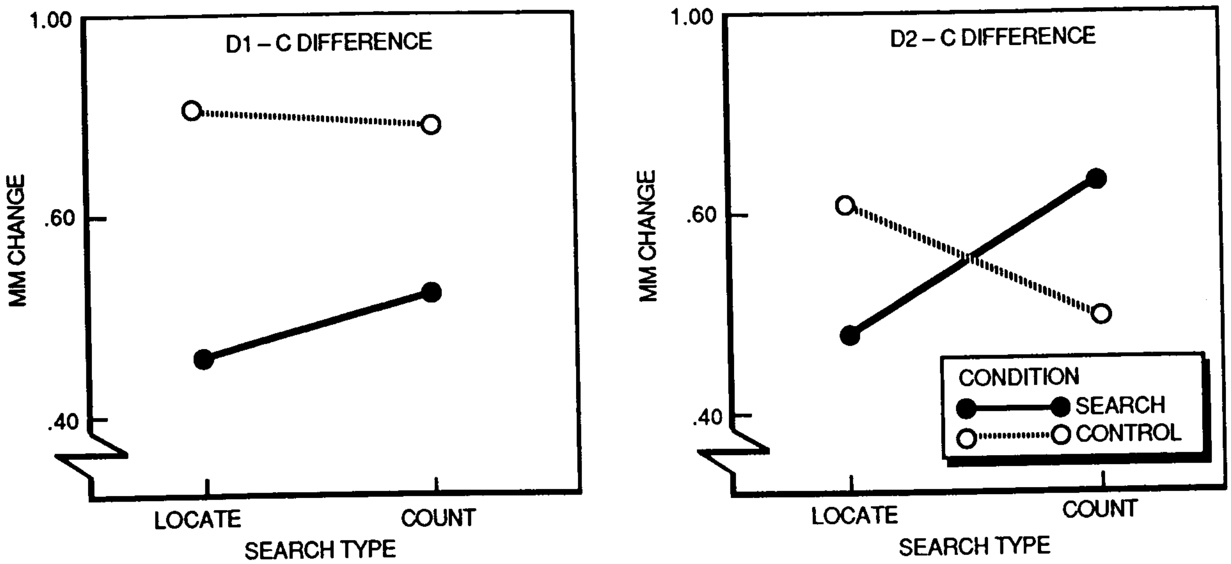

Figure 5. Search Type and Condition effects for pupillary responses (n=8).

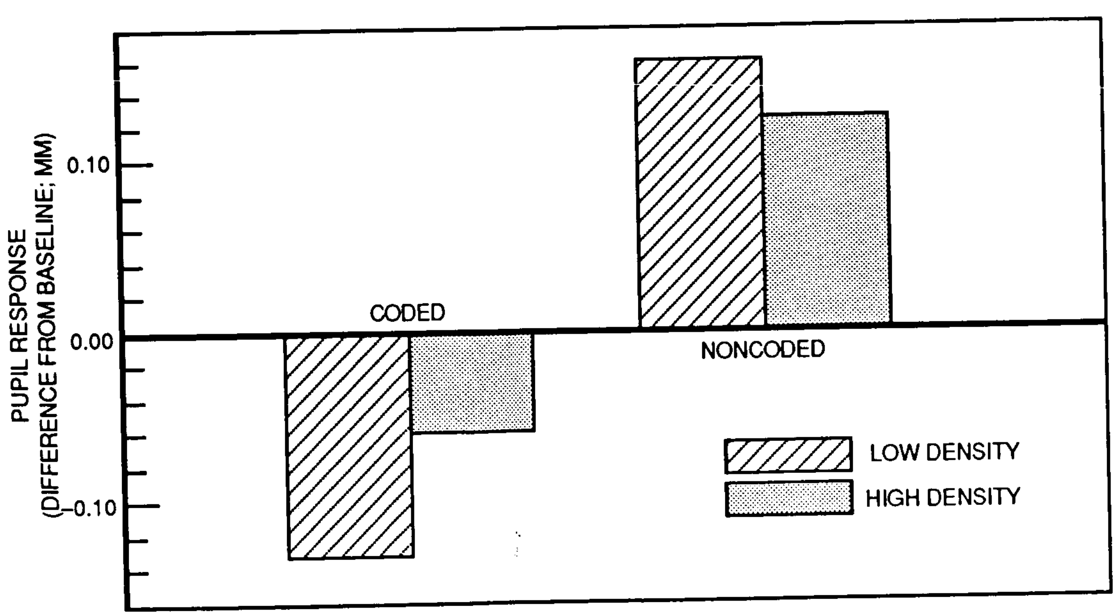

Figure 6. Density by Color Coding effect for the D2 pupillary response component (n=8).

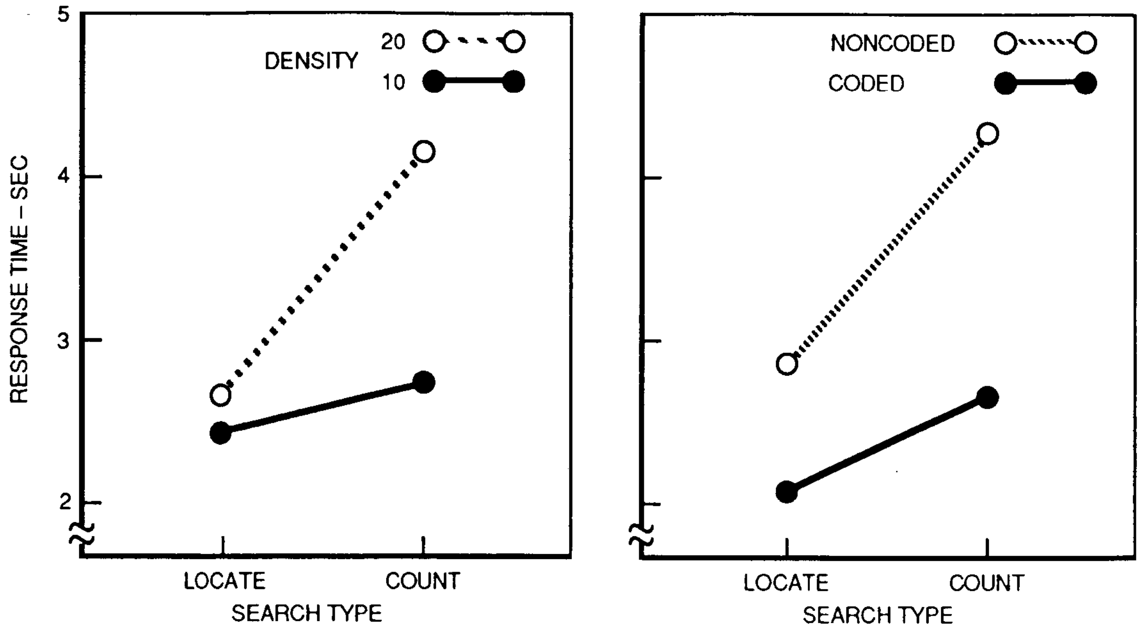

Figure 7. Effects of Color Coding, Density, and Search Type on response time (n=8).